Pitch cues for the recognition of Yes-No questions in French

Monique Vion, Annie Colas

PITCH CUES FOR THE RECOGNITION OF YES-NO QUESTIONS IN FRENCH

Monique Vion & Annie Colas *

University of Provence
CNRS UMR 6057: Parole & Langage, Psycholinguistics team
monique.vion@up.univ-mrs.fr

Résumé

L’analyse linguistique de l’intonation de la question en français montre que lorsqu’une question comporte plus de deux unités tonales, les unités intermédiaires présentent un patron d’abaissements successifs de la hauteur antérieurement à la montée finale. Ce patron constituerait une propriété phonologique de la question en français.

Le paradigme du dévoilement graduel a été utilisé pour savoir si les auditeurs peuvent reconnaître sur la base de ces indices l’énoncé d’une question. Les stimulus sont issus de phrases nominales comportant trois unités tonales.

Les résultats montrent que, bien que ce soit difficile, les auditeurs peuvent identifier l’énoncé d’une question sur la base des caractéristiques prosodiques qui précèdent le contour final, mais qu’ils ne sont en ces points que moyennement sûrs de leurs réponses.

Mots clés : français, indices mélodiques, question, reconnaissance, dévoilement graduel.

Titre courant : la reconnaissance de l’intonation de la question

Abstract

Linguistic studies of the intonation of Yes-No questions in French show that, in questions containing more than two stress groups, interrogative intonation is characterized by a sequence of lowered pitches or downstepped tones which precede the final rise.

The gating paradigm was used here to determine whether subjects listening to French NP utterances containing three stress groups could indicate whether the utterance was an statement or a question before the final rise was heard.

Although the task was difficult, findings indicate that listeners can in fact to a certain extent, recognize with mid confidence ratings, the intonation~~al~~ device of a question while they were hearing the downstepped tones preceding the final rise.

Key words: French language, melodic cues, Yes-No question, recognition, gating paradigm.

Running head: the recognition of intonation cues to questions

Speech prosody is recognized now as one of the sources of significant information for real time processing of spoken language. Although some knowledge about the role of prosodic cues exists today (see the syntheses of Cutler, Dahan, & Donselaar, 1997 and Fon, 2002), studies are devoted to this issue in psycholinguistics are still rather rare.

One of the reasons belongs to the discipline. Psycholinguistics made first the theoretical choice of studiing language processing without being aware of its specific means of expression (Segui & Ferrand, 2000). The research strategy, which consists in ignoring prosody while concentrating on the study of written materials, makes it possible to isolate the various lexical, semantic and syntactic routines available to the parser. But, following Fodor (2002), it can be objected that, every time the experimental material is presented visually, the results of the processing might be affected by the prosody projected mentally during silent reading. If this is the case, then psycholinguistics cannot "escape prosody".

Another reason of the scarcity of studies devoted to prosody in psycholinguistics is related to the difficulties, at present less technological than theoretical, that prosodic analysis meets in linguistics. As Di Cristo (2000) summarized: "[…] all measurable prosodic differences are not perceptible […] and all perceptible differences are not necessarily perceived […]; when they are, they are not bound to be perceived and when they are perceived, they do not inevitably carry-phonological distinctions ". These remarks show how much the conjunction between theoretical and methodological assets of the two disciplines can lead to a better knowledge of the role of prosody in spoken language processing.

The present study is an attempt in this direction. It falls under th e general project of appreciating the distinctive value of the differences and/or the variations revealed by prosodic analysis. It relates more precisely to the identification of the prosodic cues that allow the listener to recognize the interrogative modality in French. The search for the relevant variables for the discrimination of affirmative and interrogative utterances is not a recent one.

Some results exist for languages like Swedish, Dutch, American and Australian English (Hadding-Koch, 1961; Hadding-Koch & Studdert-Kennedy, 1964; Wales & Taylor, 1987; Swerts, Bouwhuis & Collier, 1994; Sˇaf árˇov á, & Swerts, 2004). They tend to show that the final movement of the fundamental frequency (rising vs. falling) was not the only cue taken into account by the listeners and that the totality of the contour contributes to the recognition. The present study on French takes the latter assertion as hypothesis.

Sentence modes are expressed by specific formal marks including intonation forms (i.e. the verbal mode, the subject-verb order and the intonation at the end of the sentence). Di Cristo and Hirst (1993) analyzed the prosodic regularities of various types of questions in French. In a linguistic approach, their aim was to define the prosodic structures at a phonological level[1]. The surface structure of four types of question was thus described: (1) yes/no questions ("tu veux du fromage□" Do you want some cheese?); (2) elliptic questions ("et du dessert□" and dessert?); (3) questions which propose a choice ("tu veux du fromage ou du dessert?" Do you want some cheese or dessert?) and (4) questions with a left-dislocation ("du dessert, tu en veux aussi? " dessert, you want some too?). The corpus consisted of radio recordings of talks and short simulated dialogues. The analysis showed two types of final movement of the fundamental frequency. Whereas the yes-no questions as well as the first term of the alternative questions ("tu veux du fromage…" Do you want some cheese…) and the rhematic part of the questions which comprise a left-dislocation ("tu en veux aussi?" you want some too?) showed a final rise, the elliptic questions as well as the second term of the alternative

---

[1]For a synthetic presentation of the theory of intonation developed by the authors (Hirst and Di Cristo, 1984; Hirst, 1988; Hirst and Di Cristo, 1998; Hirst, Di Cristo & Espesser, 2000). One can refer to Rossi (1999) or Di Cristo (2000).

questions and the thematic part of the questions with left-dislocation (of the dessert…), showed a falling final contour. Moreover whatever the type of the question, when the intonation units [2] included more than two Tonal Units (TU) [3], the overall description of pitch variations showed the presence of successively lowered pitches for the intermediate TUs. When the question was made of one or two TUs, the pattern looked like the declarative one except for the final rise (see Table 1, the example on the left-hand side). When the question included more than two TUs the pattern of the intermediate TU was characterized by a sequence of lowered pitches (see Table 1, the example on the right). According to the authors, the pattern of downstepped tones for the intermediate TUs constitutes a phonological property of questions in French and not just a feature related to the anticipation of the final rise.

	Two Tonal Units	Three Tonal Units
Statement	L'aMI de JEAN [⇑ > ⇓]	L'aMI du voiSIN de JEAN [⇑ ↓ ≠ > ⇓]
Question	L'aMI de JEAN [⇑ > ⇑]	L'aMI du voiSIN de JEAN [⇑ > > ⇑]

Table 1: Pitch variations within the tonal units (statement vs. question)(from Di Cristo, 1998).

---

[2] Speech prosody is made up of units of varied size, called Intonation Units (IU) that are regarded as the prosodic broadest unit. In the theory developed by Hirst & al. (2000), the IU is a concatenation of Tonal Units (see note 3), limited by a specific tone boundary (conclusive or not conclusive; the tone is the height unit of the intonation).

[3] The Tonal Unit (or stress group) is a minimal unit of synchronization of the tonal segments and phonemes of the speech flow.

Caption: The accented part is written in capitals; ↓: lower; ↑: higher; >: downstep; ⇑ :top (maximum height): ⇓ : bottom (notation INTSINT, Di Cristo, Hirst, Boudouresque & Louis, 2002).

Many verbal exchanges in ordinary life show cases in which intonation alone may carry an indication of the sentence mode and the speaker's intention (" tu as bien dormi□" Did you sleep well?). The goal of our experiment was to show that the lowered pitches which precede the final rise when intonation units comprise more than two tonal units, are relevant features to identify a question. In other words, these lowered pitches are cues which allow the listener to recognize the sentence mode.

The use of the gating paradigm seemed a good way to try to show this. The technique is relevant for a lot of studies which try to determine the amount of acoustico-phonetic information needed to identify a stimulus, or what role is played by the phonetic, lexical and contextual variables during identification (Grosjean, 1996). Up to now the gating paradigm has rarely been used to study the role of prosodic cues in spoken sentence processing (Grosjean, 1980a; Grosjean and Hirt, 1996). Its most frequent use is in the domain of lexical identification. In the version proposed by Grosjean (1980 a and b), a series of stimuli including incremented portions of an element of spoken language (syllable, word, phrase or sentence) is generated from the onset of the signal. The first gate is very short and the last gate includes the totality of the linguistic element. The increment size of the gates can be expressed in milliseconds or as a percentage of the target element. The listener is asked: (1) to identify the target element and (2) to indicate how confident s/he feels about her/is answer at each gate. Three dependant variables can be derived from the responses: 1) the isolation point (i.e. the size of the target segment - duration or % of the target element- needed for a correct identification; 2) the confidence ratings at any point of the target element (in particular at the isolation point and at the end of the target element); and 3) the candidates proposed before the identification was correct.

Noun Phrase (NP) utterances (also called non-verbal utterances) are a kind of sentence where intonation plays a decisive role in the formal marking of sentence mode. They are the best experimental material for the study of the perception of prosody (Swerts, Bouwhuis & Collier, 1994). NP utterances, found sometimes in the form of fixed formulas (" au diable l'avarice! " to hell with miserness! ," gloire à Dieu au plus haut des cieux et paix sur la terre aux hommes de bonne volonté. " glory to God in heaven and peace on earth to men of good will.), are very frequent in spontaneous spoken language ("très bien ton exposé " good speech, well done.). They are considered the fruit of an immediate reaction of the speakers, ordered by the dynamics of ~~the~~ communication.

From a grammatical point of view, although being sentences without a verb, they are complete sentences insofar as they are composed of two terms, a subject (your speech) and a predicate (good), connected by the speaker (Le Goffic, 1993). The syntactic means used to assert using these sentences are prosody and word order. As one can see in the examples above, the absence of a verb does not prevent the sentence mode (affirmative or exclamative generally).

From a pragmatic point of view, interrogative NP utterances seem rather to elicit a yes/no response. It can be a question or a request for confirmation (a: "je me suis inscrit en troisième année de droit " I was registered for third grade course in law school, b: " en troisième année de droit ? " for third grade course in law school?) or of a dubitative resumption of an assertion (a: " tu as réussi ton examen" you passed your exam, b: "j'ai réussi mon examen ? " I passed my exam?).

NP utterances can consist of either a single word ("imbécile□" stupid!) or a sequence of words the length of which is only limited by the immediate storage capacity (a: "tu fais quoi, toi, dans la vie□" what do you do for a living? b: "adjoint du directeur du centre de recherche géologique de l'institut régional de diffusion de la science de l'académie de Montpellier" assistant of the director of the geological research center of the regional institute of diffusion of science of the academy of Montpellier). From a methodological point of view, this property enables the experimenter to build stimuli which can comprise a variable number of TUs.

In the NP utterances selected for this study the nominal term represents a predicate of existence (existential nominal sentences). The nominal group, a single component, incorporates what could be developed in a sentence.

METHOD

Materials

Sentences

The NP sentences which were used to generate the stimuli for the experimental trials and for the practice session consisted of three tonal units. Each tonal unit included three syllables. In order to preserve an uninterrupted pitch curve, all the syllables included a voiced Consonant and a Vowel (see the sentences in appendix 1). The sentences came from the materials worked out at the time of pilot studies (Vion & Colas, 2002)[4].

---

[4]We warmly thank Isabelle Soulier (graduate student at the time) for her help with the pilot experiments, and also Albert Di Cristo and Daniel Hirst ("Prosody and formal representation of language" team of the Laboratory Parole et Langage) for their often requested expertise.

Each sentence was tape-recorded twice by a female speaker (one of the authors): one with the intonation of a statement and the second with the intonation of a question. The recordings were submitted to an expert in order to make sure: (1) that the way of pronouncing the affirmative and interrogative sentences was homogeneous and(2), that the intonation of the sentences was in conformity with the pattern described in the literature. The statement is decribed as (Figure 1a): " *a basic pitch pattern (i.e. an Intonation Unit) ending on a final low pitch (such as a statement) contains a rising pitch movement (from low to high) at the end of each stress-group, except the last which is pronounced with a falling pitch movement from the mid (or lowered high) to low*." Di Cristo (1998), p. 201. As for the question (Figure 1b): "*an initial peak situated on the last syllable of the first stress-group followed by a declining pitch until the final rise*." Di Cristo (1998), p. 203.

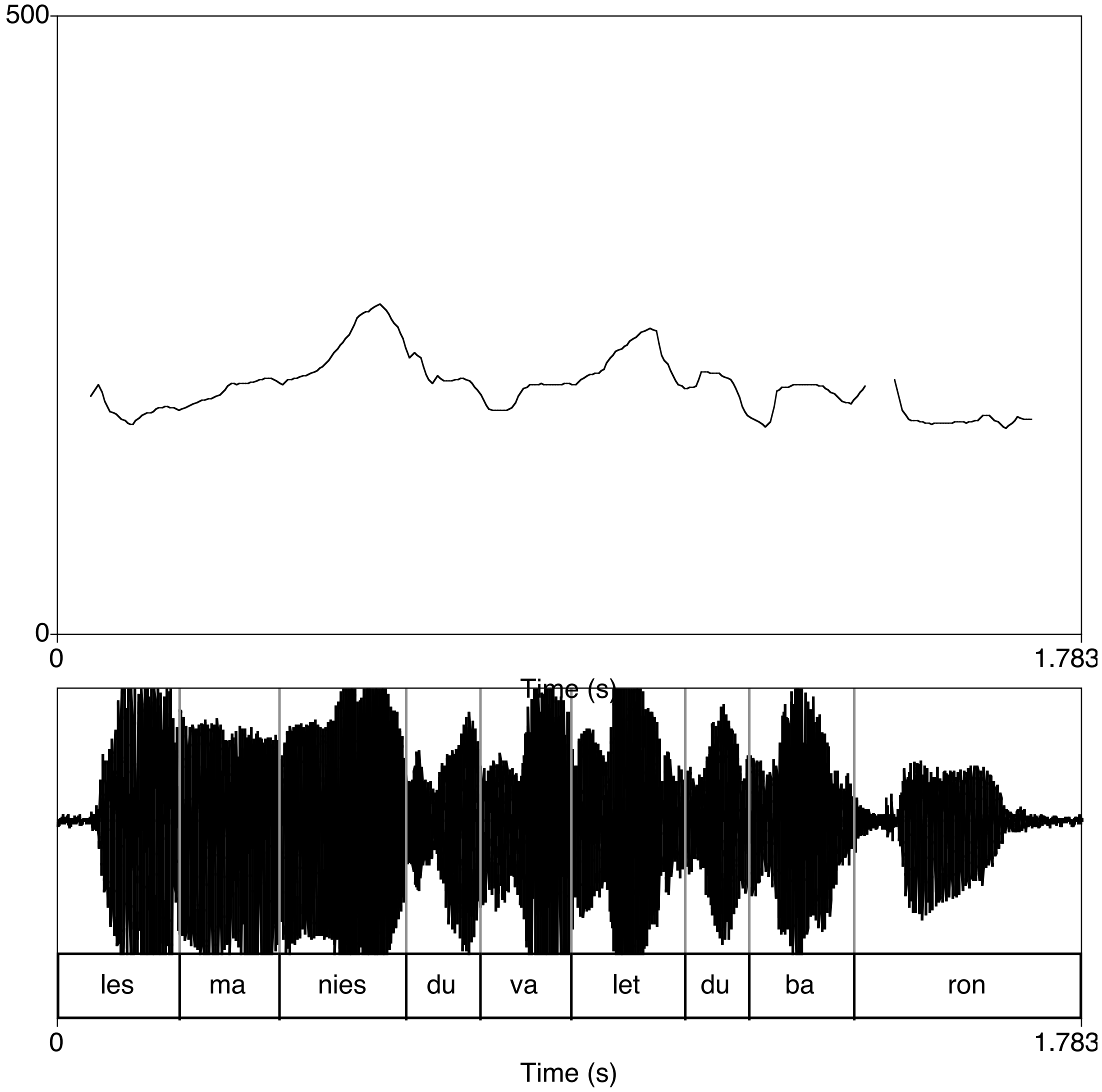

fig.1 (a)

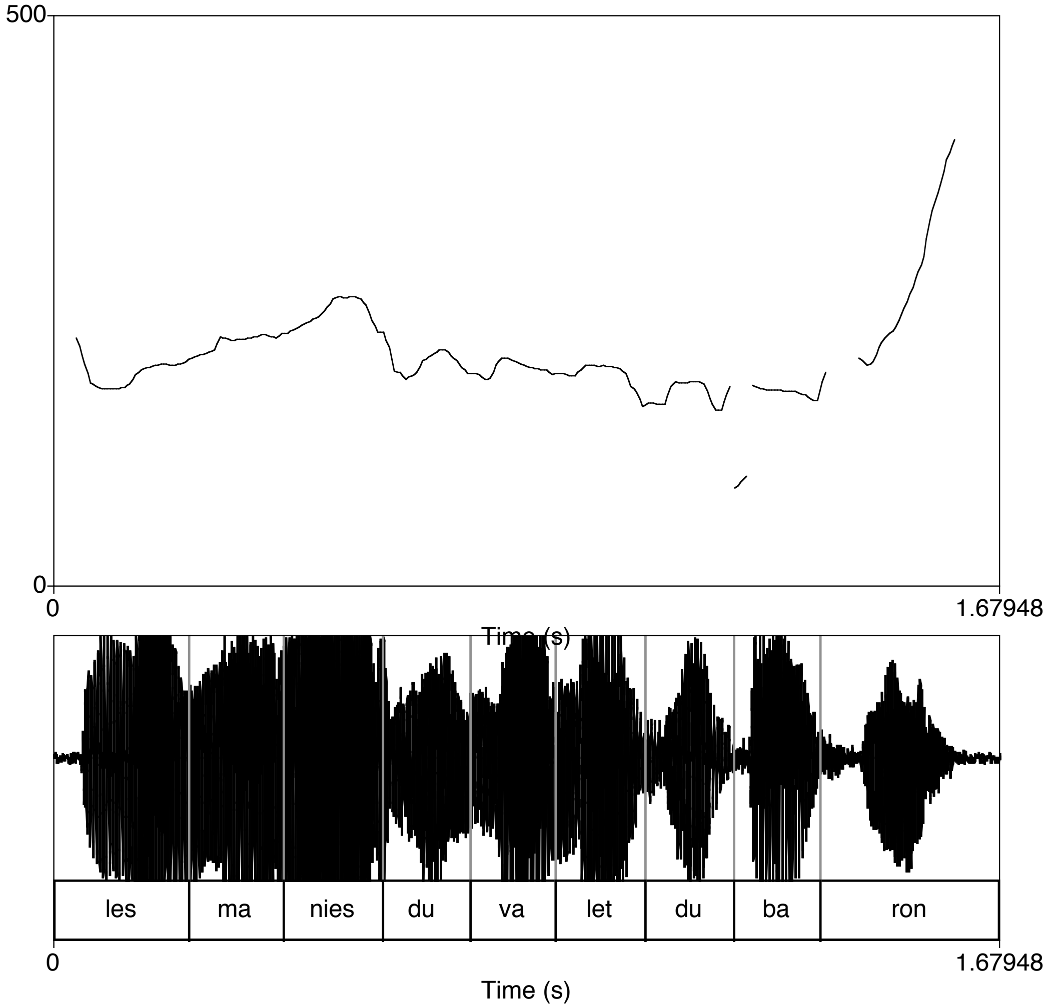

fig.1 (b)

Figure1: Acoustic analysis of the sentence " les manies du valet du baron" (the manias of the servant of the baron), uttered as an statement (a), and as a question (b).

High: fundamental frequency variations

Low: acoustic signal

Stimulus

For each tape-recorded utterance a set of eleven stimuli (gates) was generated. Tape recordings were converted to files and gated using the SoundEdit 16 system. The first stimulus (s1) included the first TU made of three syllables. In addition to the three syllables of s1, the five following gates included each, an incremented fragment of the syllables, which constitute the second TU. The increment size of the gates was 20%. Stimuli s2, s3, s4, s5 and s6 included respectively 20%, 40%, 60%, 80% and 100% of the duration of the second TU. In addition to the six syllables of the first two TUs, the last five gates included an incremented fragment of the syllables that constitute the third TU. The increment size of the gates was 20%: stimuli s7, s8, s9, s10 and s11 respectively included 20%, 40%, 60%, 80% and 100% of the duration of the third TU. It should be noted that for the Q series, s9 included the totality of the declining pitch and that the NP utterance was complete in the last gate (s11). Comparison of pairs of S and Q series showed that, whatever the contents of the sentences, the pairs s5 and s6 (TU1+80% of TU2 and TU1+100% of TU2) as well as the pairs s10 and s11 (TU1+TU2+80% of TU3 and TU1+TU2+TU3) presented clearly audible pitch differences.

Items for the practice session

To familiarize the participants with the task, a list of three items (three sets of eleven stimuli generated as described above) was presented so that the participants could hear the intonation of the statement, of the question, and of the implication. For all the participants the order of presentation of the three items was: statement – implication - question.

Items for the experimental trials

Two lists were created for the items test so that the contents of a sentence whould be heard in one list associated with the intonation of the statement and in the other list with the intonation of the question. In a list, four contents were thus presented in the form of a statement and four in the form of a question.

Procedure

The experiment was controlled by a computer. The participants were tested individually. Before beginning the experiment, the experimenter invited the listener to read aloud the instruction progressively displayed on the computer screen. The experimenter presented them step by step while clicking on the mouse and commented on it if needed (see the instructions in appendix 2). Then the experimenter invited the listener to put on the headphones.

The listener was first introduced to the three items of the practice session, and then to the eight items of the experiment itself. For each gate, the program recorded an answer ("S" or "Q") and its time latency, the confidence rating and its time latency. The experiment lasted on average thirty minutes.

The practice and the experimental trials had the same structure. A stimulus was first played. Immediately afterwards, the two possible versions of the sentence appeared one below the other on the top of the screen (the statement was preceded by the label S and included a final stop; the question was preceded by the label Q and included a question mark). A text beneath invited the participant to answer ("S" or "Q") by pressing one of the two keys (labeled "S" or "Q") on the keyboard.

Once the answer was recorded, a new text on the screen invited the participant to indicate by pressing one of the 9 keys of the number pad how confident he felt about his/her answer (with "1" meaning: "not certain at all" and "9": "very certain "). A third text on the screen then invited the participant to begin the next trial by pressing the space bar. At the end of the last trial of a series, a text on the screen indicated the end of the current item and invited the participant to begin a new item by pressing the space bar.

Factors

Two independent variables were manipulated: the *intonation of the sentence* (with two categories: statement (IS), question (IQ)) and the *the amount of information input or size of the stimulus* (with 11 categories: from s1 - first tonal unit - to s11 - complete sentence). The *identity of the speaker* (female voice), the *number of tonal units* (three units) and the *size of the tonal units* (three syllables) were constants. There were three variables of control: the *list of items,* (with two categories) which controlled a possible effect related to the contents of the sentences; the *display on the screen of the written sentences* (with two categories: above – below) which counterbalanced a possible order effect; and the*display of the response keys* (with two categories: Left - Right) which counterbalanced a possible order effect.

Data collection design

Each participant was confronted with both *intonation patterns* and listened to each content with only one of the possible patterns. This implied the constitution of two groups of participants so that each content was presented with both intonation patterns in the experiment. Each participant was confronted with only one display *of the written sentences.* This again implied the constitution of two more groups. Each participant was confronted with only one response *key display,* which finally implied the constitution of two more groups. The experiment thus required the creation of eight groups of participants (resulting from taking into account all combinations of the three control variables each with two categories: list 2 X screen 2 X keyboard 2) with four participants each.

The programs written for the experiment (software PsyScope□Cohen, Mac Whinney, Flatt and Provost, 1993) differed by the list and the display of the written sentences on the computer screen. For each one, the statements and questions were presented in an order drawn randomly for all eight items. Moreover, for a given program, half of the participants had the key labeled "S" on the left of the keyboard and the key labeled "Q" on the right. For the other half the display of the keys was reversed.

Participants

Thirty-two young adult native speakers of French, students at the University of Provence took part in the experiment.

Predictions

From a descriptive point of view, the prosodic structure of the affirmative and interrogative sentences in French appears different only from the second TU. For the sentences used in the present experiment that included three TUs, the assumption was that a listener could start to differentiate the pattern of the question from that of the assertion as he/she hears the lowered pitches of the second TU. Predictions thus relate to the evolution of the identification responses as the amount of discriminating prosodic information increases. If the responses given by the listeners are grounded on the processing of prosodic cues, then the number of correct identifications should increase as the size of the part of the signal carrying the lowered pitches increases.

RESULTS

Each participant gave 88 responses of identification (8 items X 11 gates) and expressed as many confidence ratings. 2816 responses of identification (32 participants X 88) and 2816 confidence ratings were thus available for the analysis. Moreover, the time latency of each answer was also available.

Overall results

The data were first examined from the point of view of the time taken to identify a stimulus whatever the correctness of the response. An analysis of variance was conducted with the mean response time as dependent variable. The analysis design was: 2 (*Intonation)* x 11 (*Size of stimulus*). The analysis yielded a significant effect of the size of stimulus and an interaction effect between this factor and the Intonation factor (figure 2).

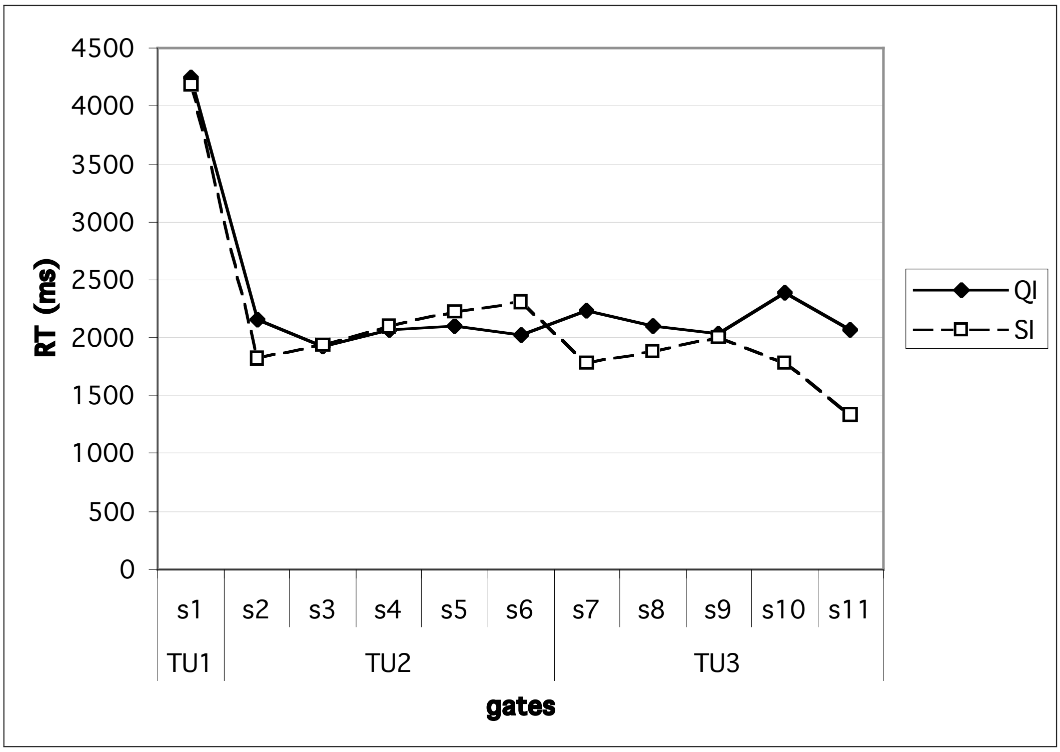

fig.2

Figure 2: Evolution of the mean latencies of the responses of identification during tests.

Caption:

SI: declarative intonation; QI: interrogative intonation; s1: first tonal unit (TU1); s2: TU1+20% of the second tonal unit (TU2); s3: TU1+40% of (TU 2); s4: TU1+60% of (TU2); s5: TU1+80% of (TU2); s6: TU1+TU2; s7: TU1+TU2+20% of the third tonal unit (TU3); s8: TU1+TU2+40% of TU3; s9: TU1+TU2+60% of TU3; s10: TU1+TU2+80% of TU3; s11: TU1+TU2+TU3.

Globally, the mean response time, which was about 4 seconds for the first gate TU1), decreased to around 2 seconds for the second gate (TU1+ 20% of TU2) and stabilized around this value there after [$\underline{F}$ (10,310) = 32.138; $\underline{p} < 0.00001$]. A pairwise comparison using the Neuman-Keuls test confirmed (at 0.01) that only the RT associated with the first gate (s1) differed significantly from all the others. With regard to the interaction effect [$\underline{F}$ (10,310) = 4.377; $\underline{p} < 0.00001$], for each gate in a series, the comparison of the RT obtained for the stimuli A with the RT obtained for the stimuli Q showed that they differ significantly only for the gate s10 (TU1 + TU2 + 80% of the TU3) and s11: the RTs for the stimuli S were shorter than the RTs for the stimuli Q (1.5 seconds vs. 2.2 seconds).

The data were then examined considering the correctness of the answers. A dependent variable was derived from the answers "S" or "Q". The answer "S" given for the stimuli S and the answers "Q " given for the stimuli Q were regarded as "correct answers " and coded “1”; in the other case the answers were regarded as incorrect and coded “0”. A variance analysis according to the design described above and carried out with this dependent variable (the extent of which varied from 0 to 4) showed a significant effect of each factor and an interaction effect between them.

Globally, gates S triggered more correct responses than gates Q (on average 3.3 answers out of 4 vs. 1.9) [$\underline{F}$ (1,310) = 42.912; $\underline{p} < 0.00001$]. And on average, the correct responses, which equaled 2.5 from the first to the ninth gate, equaled 2.9 for gate s10 and 3.8 for the last gate (s11) [$\underline{F}$ (10,310) = 29.335; $\underline{p} < 0.00001$]. A pairwise comparison using the Neuman-Keuls test confirmed (at 0.01), that only the responses given for the gates s10 and s11 differed significantly from those given for the other stimuli of the gates. The examination of figure 3 which presents the interaction between *the Intonation* and the *Size of stimulus* factors [$\underline{F}$

(10,310) = 19.638; p < 0.00001] showed that the difference between the correct responses to gates S and to gates Q was reduced as the size of stimulus concerning the TU2 increased.

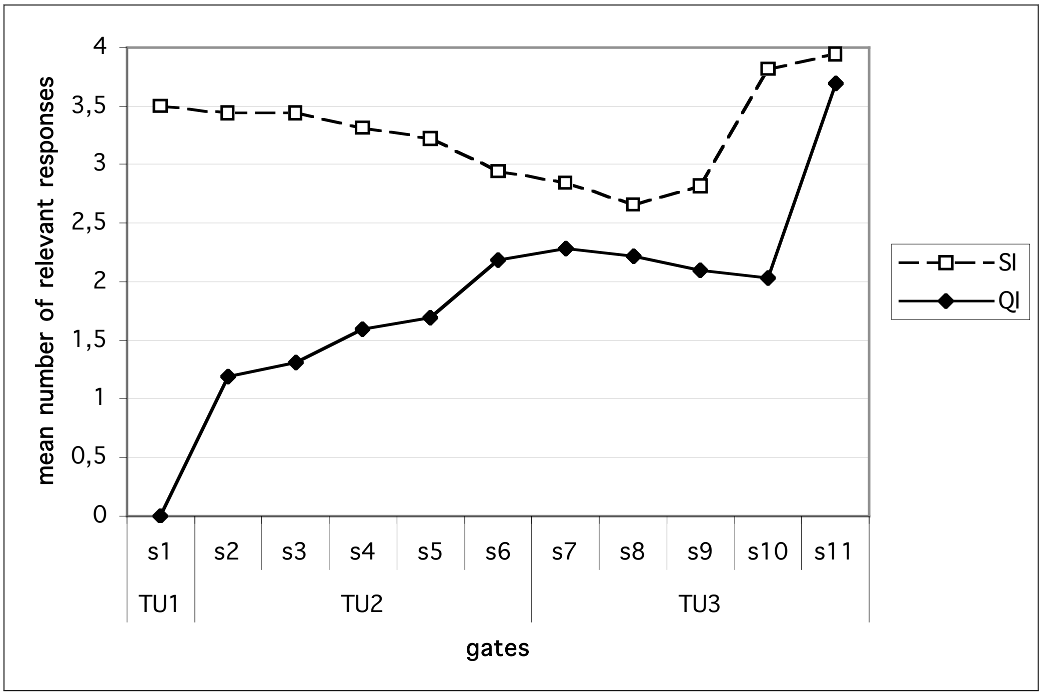

fig. 3

Figure 3: Evolution of the mean number of correct responses during tests.

Caption:

SI: declarative intonation; QI: interrogative intonation; s1: first tonal unit (TU1); s2: TU1+20% of the second tonal unit (TU2); s3: TU1+40% of (TU 2); s4: TU1+60% of (TU2); s5: TU1+80% of (TU2); s6: TU1+TU2; s7: TU1+TU2+20% of the third tonal unit (TU3); s8: TU1+TU2+40% of TU3; s9: TU1+TU2+60% of TU3; s10: TU1+TU2+80% of TU3; s11: TU1+TU2+TU3.

The comparison of the difference for each gate showed that the difference between the correct responses S and Q remained stable until the fifth gate. The advantage which gates S to profit from for a time can be interpreted as a response bias, resulting from the application of a strategy which would consist in prefering to answer "S" rather than "Q"from the beginning of a series of stimuli while referring to the frequency of the two types of statements in language use, then to eventually modify this attitude for the upcoming gate. The examination of the responses for the first gate (TU1) showed indeed that in the absence of any discriminating prosodic information, in 67.2 % of the cases, the participants responded "S". For the sixth gate (TU1+TU2) the difference between the answers reduced: the correct answers to stimuli S decreased slightly whereas those to the stimuli Q increased more. This new variation then remained stable until the ninth gate. For the tenth gate (TU1+TU2 + 80% of TU3), the number of correct responses to stimuli S increases, whereas that to the stimuli Q remained stable. It was only for the eleventh gate (TU1+TU2+TU3) that the number of correct responses to the stimuli Q increased in turn. It should be noted that when the whole sentence was presented, incorrect responses still existed which represented 1.6 % of the total of the responses to stimuli S and 7.8 % of the total of the responses to the stimuli Q.

To summarize, the examination of the overall data brought three points to light: (1) the presence of a possible strategic bias in the responses (the participants showed a tendency to answer rather "S" than "Q" from the start); (2) a regression of this tendency when the stimulus was composed of TU1+TU2; and (3) the importance (visible as well on the responses as on the time it took to give them) of the acoustic characteristics available in all end of the NP utterance (in the last 40 % of the third TU).

Analysis of the patterns of responses

To confirm the interpretation of the progressive reduction of the difference between the responses given for gates S and gates Q by the fact that the participants noticed the discriminant prosodic cues, it was necessary to study not only the number of correct responses given by the group of participants, but also for each participant to study the distribution of the responses within each series. It was thus necessary to examine, for each participant, eight patterns comprising eleven responses each.

If, for a series of stimuli, a listener decided to answer from the beginning "S" and to go on until proof found, within the increasing amount of input, that this answer was not appropriate, and if the series he/she was confronted with was ~~a~~ S series , then all the answers would be correct. The resulting pattern is a succession of eleven: "1_s". If the series was a Q series and if the listener did not discover in the increasing amount of input contrary cues to his/her initial answer "S", then the response pattern would be a succession of eleven:"0_s", it would be the opposite for a listener who decided from the begining to answer "Q". Finally if it seemed necessary for the participant to change his/her answer while hearing a series, then the pattern would be a compound of a succession of zeroes and ones (heterogeneous pattern).

The examination of the 256 patterns (8 items X 32 participants) (Figure 4) showed that the participants did not modify their initial response while hearing the series of stimulus in 38.3% of the cases. As we conjectured in the light of the overall results, the dominant strategy consisted for the listeners in answering systematically "S" (28.9%) rather than "Q " (9.4%).

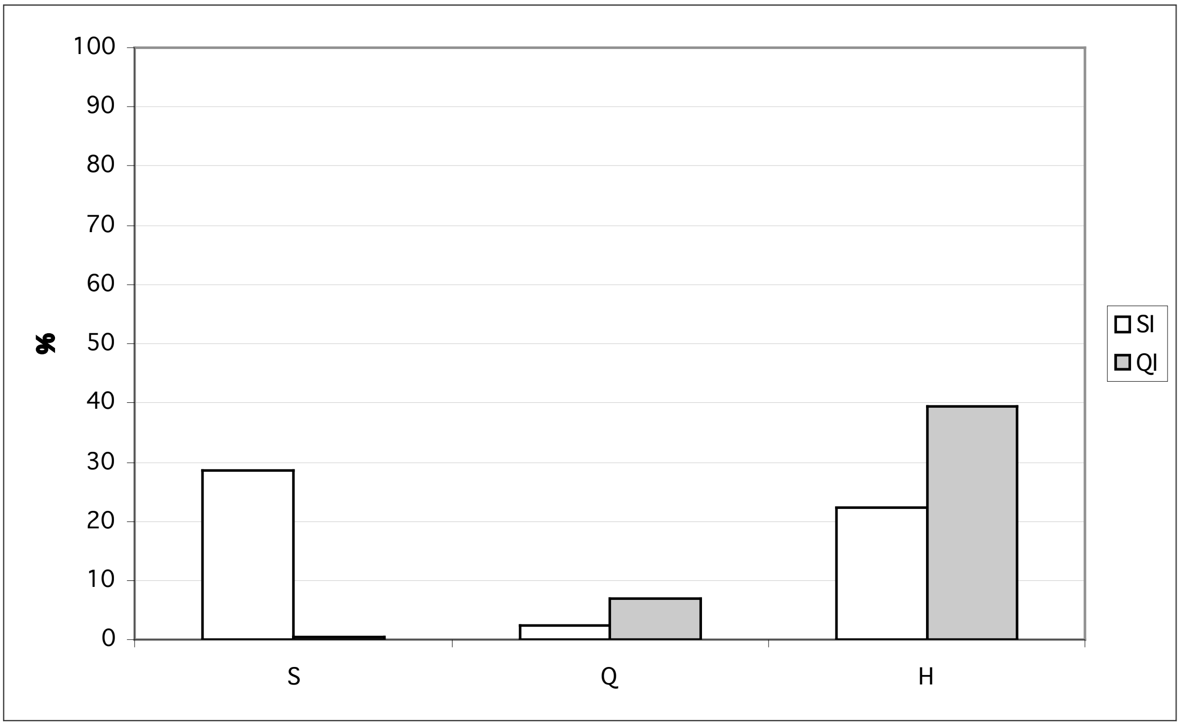

fig.4

Figure 4: Distribution of the response patterns according to the intonation of the stimulus.

Caption: SI: declarative intonation; QI: interrogative intonation; S: systematic answer "S "; Q: systematic answer "Q "; H: heterogeneous answers.

This result underlines the difficulty of the participants in focussing their attention on the prosodic characteristics of the stimuli. However, none of the 32 participants ever deviated from his/her initial choice. The heterogeneous pattern (61.7 % of the patterns), which attest a non-systematic answer to the stimuli, is found in the performance of each participant.

Only an examination of the heterogeneous patterns can make it possible to provide arguments in favor of the assumption of the listener's capacity to identify the distinctive features of the prosodic structure before hearing the final rise. To do that, a new dependent variable was derived from the 158 heterogeneous patterns.

Heterogeneous patterns can be the result either of a change of answer during the series of stimuli - not-founded on perceptive cues - or of an active processing of the characteristics of the intonation they heard step by step. In both cases, two types of heterogeneous patterns can be observed. First the pattern presents itself as an irregular sequence, made of "0 " and "1 " ("irregular pattern”). Second the pattern presents itself as a regular one: it begins with "0" and goes on until the occurrence of "1 " which is maintained until the end ("regular pattern"). Figure 5 presents the distribution of both types of patterns according to the intonation of the stimuli. The irregular patterns, although non-dominant, represented about half of the heterogeneous patterns (48.8 %) and concerned essentially the S series. The irregular patterns that were dominant (53.16 % of the patterns) concerned especially the Q series.

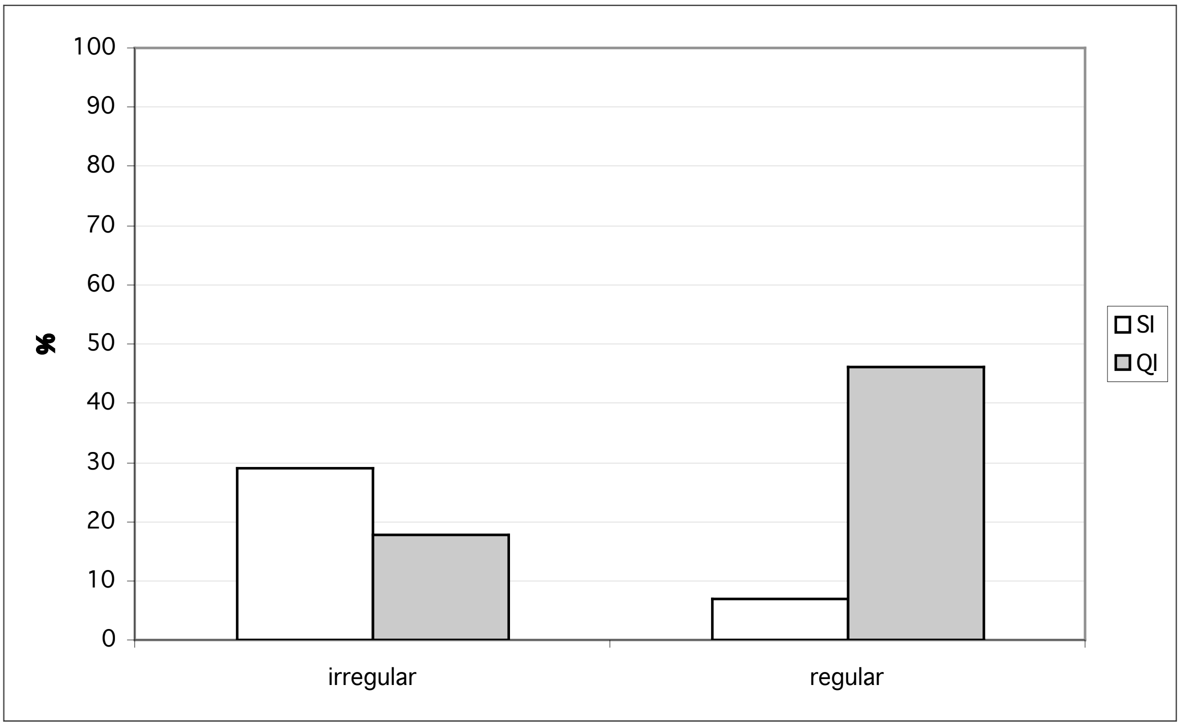

fig.5

Figure 5: Heterogeneous response patterns: distribution according to the intonation of the stimulus.

Caption: SI: declarative intonation; QI: interrogative intonation; irregular : irregular heterogeneous pattern; regular : regular heterogeneous pattern.

Moreover it is interesting to note that among the 73 regular patterns observed only 10 participants show regular patterns in response to the S series, whereas 29 participants provide them in response to the Q series. So that 90.6 % of the participants show a regular pattern at least for one Q series. In a more detailled description: four participants show this pattern for only one Q series, as opposed to twenty of them who show it for two or three Q series and five participants who show it for the four Q series.

The presence of regular patterns in response in to the Q series by a vast majority of listeners allowed us to continue the analysis and to determine which was the quantity of entering prosodic information, which made it possible to recognize the sentence mode. Figure 6 presents the distribution of the regular patterns according to the point at which the first correct answer was given in a series.

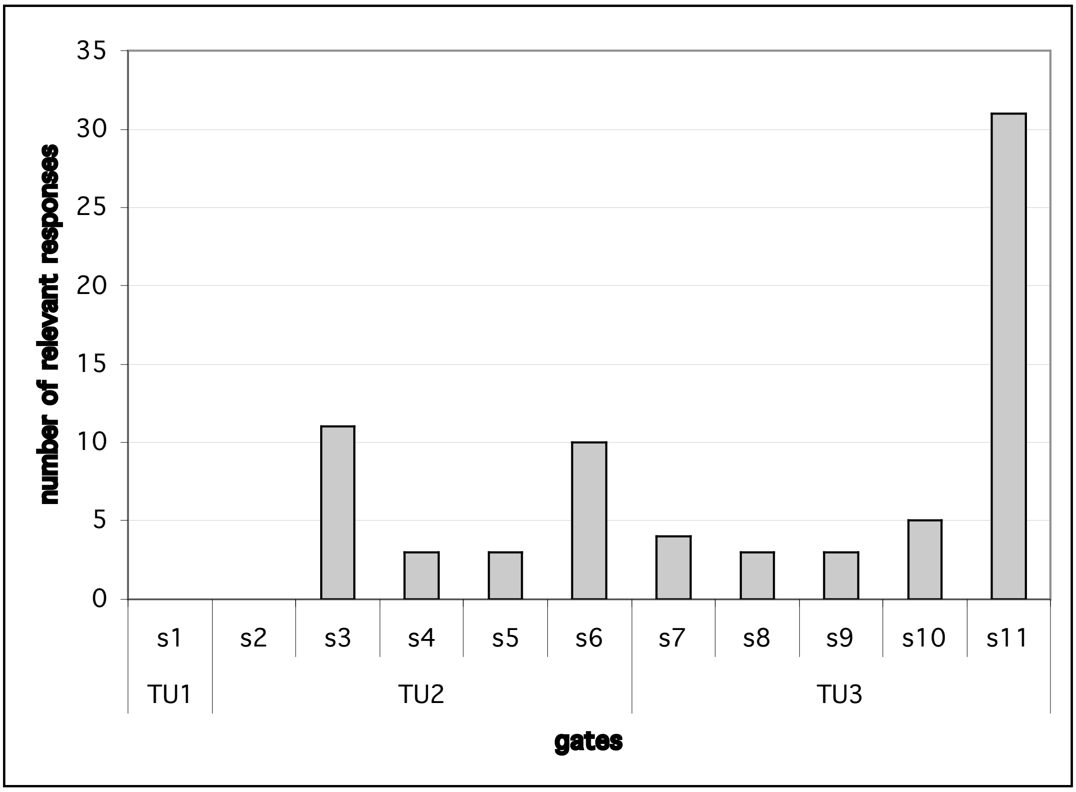

fig.6

Figure 6: Regular patterns: distribution according to the data point of stimulus Q where the first relevant answer was given.

Caption:

S1: first tonal unit (TU1); s2: TU1+20% of the second tonal unit (TU2); s3: TU1+40% of (TU 2); s4: TU1+60% of (TU2); s5: TU1+80% of (TU2); s6: TU1+TU2; s7: TU1+TU2+20% of the third tonal unit (TU3); s8: TU1+TU2+40% of TU3; s9: TU1+TU2+60% of TU3; s10: TU1+TU2+80% of TU3; s11: TU1+TU2+TU3.

Among the 73 regular patterns, 31 were such that the first correct answer was given at the last gate of the series. This result underlines the importance of the end of the rising movement of fundamental frequency in the understanding of the sentence mode. The distribution of the other 42 patterns included two modes. For about half of the remaining patterns the hearing of the gates s3 and s6 was the occasion to give the first correct answer of a series. Whereas the correct answer given to the stimulus s6 (TU1+TU2) can be explained by the presence of perceptible prosodic differences between gates s6 S and s6 Q , the correct responses given to gate s3 cannot be explained in this way. Rather the answer seems to be a matter of decision-making In the absence of valid cues. The listener could decide to change the nature of his/her answer “just to press another key”.

To go further in the analysis of the regular patterns, a unique pattern was associated with each of the 29 participants who at least once gave a regular pattern in response to a series of stimuli Q. For four of the participants the new pattern corresponded to the only regular pattern given and for the other twenty-five participants, to a pattern resulting from the calculation of the average of the correct responses given in each data point. The analysis of variance carried out with the factor *Size of stimulus* and the frequency of the relevant answer in each data point for the dependent variable showed a significant effect for this factor [$\underline{F}$ (10,280) = 42.244; $\underline{p}$ < .00001]. A pairwise comparison using the Neuman-Keuls test showed (at 0. 01) a set of significant differences between pairs of stimuli (figure 7).

The comparison confirmed the decisive role of the end part of the ascending movement of F0 provided by the last gate. The frequency of the correct answer given for the gate s11 was very high (0.97) and differed significantly from the frequencies observed for all the other gates. Segmental information (i.e. the content of the sentence) was completely available in gate s10, so that in gate s11 only pitch information was added. It should be noted also that the frequency of the correct answer equaled 0.61 for the gate s10 and exceeded 0.50 only starting from the ninth stimulus (0.51), i.e. after 60% of TU3 was available. At this point, the declining pitch preceding the final rise was completely available.

The frequency of the correct responses observed for gate s6 (TU1+TU2: 0.37), was higher and differed significantly from that observed for gates s4 and s3. The frequency observed for gate s5 (0.22), which represents TU1+80 % of TU2, differed significantly from that observed for the gates s1 and s2. Insofar as the pairs of stimulus S/Q in s5 and s6 present definitely perceptible pitch differences, these last two results are compatible with the hypothesis of an increase in the frequency of the correct answer related to the detection of melodic characteristics suitable for the prosodic organization of the UT2.

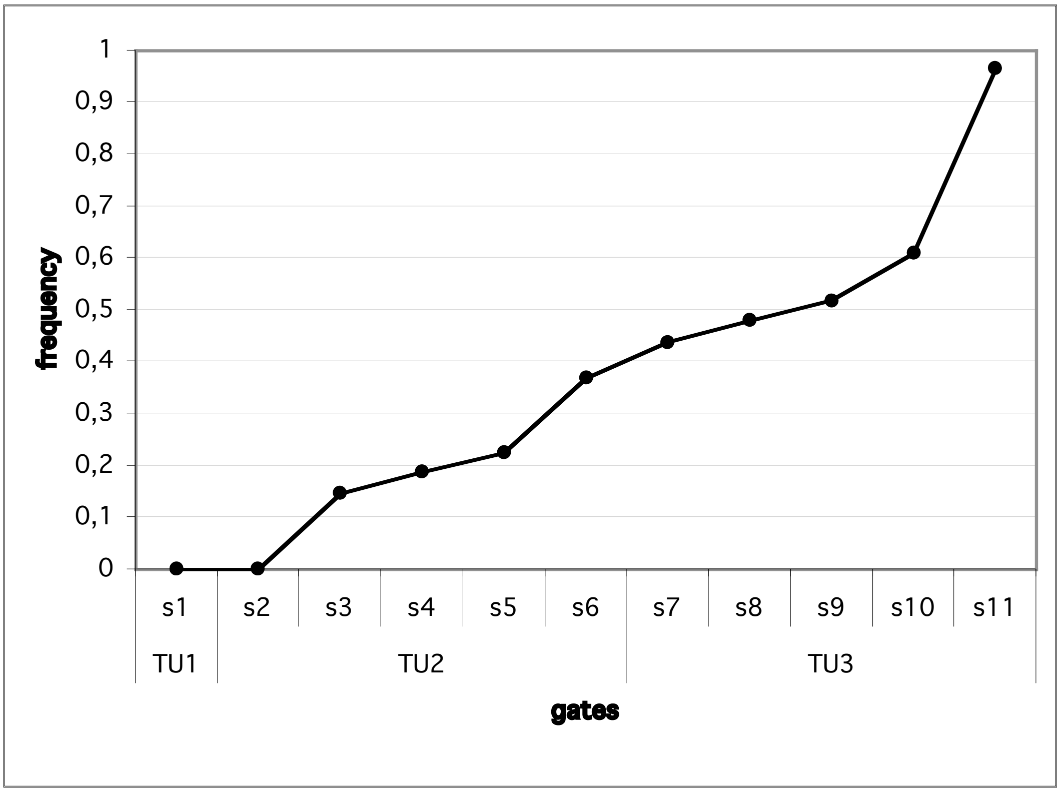

Figure 7: Regular patterns for the series of stimulus Q: evolution of the frequency of the relevant answer during tests.

Caption:

S1: first tonal unit (TU1); s2: TU1+20% of the second tonal unit (TU2); s3: TU1+40% of (TU 2); s4: TU1+60% of (TU2); s5: TU1+80% of (TU2); s6: TU1+TU2; s7: TU1+TU2+20% of the third tonal unit (TU3); s8: TU1+TU2+40% of TU3; s9: TU1+TU2+60% of TU3; s10: TU1+TU2+80% of TU3; s11: TU1+TU2+TU3.

The data collected also make it possible to analyze the time it takes to recognize each gate Q as well as the confidence ratings associated with each answer. The analysis of the mean response time for the regular patterns reproduced the same pattern of results as described for the overall data. The analysis of the mean confidence ratings for the regular patterns indicated a significant effect of the amount of entering information [$\underline{F}$ (10,280) = 64.128; $\underline{p} < 0.00001$] (figure 8).

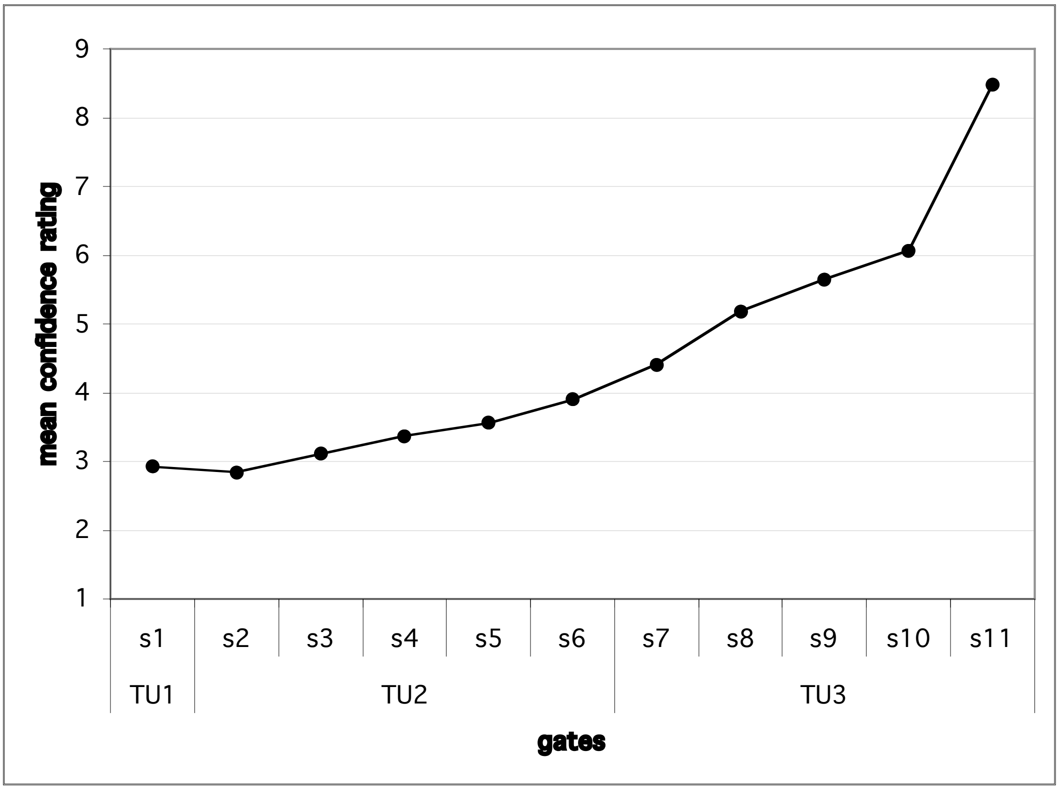

Figure 8: Regular patterns for the series of stimulus Q: evolution of the mean degree of confidence during tests: 1 = "not certain at all" and 9 = "very certain"; N = 26 participants).

Caption:

S1: first tonal unit (TU1); s2: TU1+20% of the second tonal unit (TU2); s3: TU1+40% of (TU 2); s4: TU1+60% of (TU2); s5: TU1+80% of (TU2); s6: TU1+TU2; s7: TU1+TU2+20% of the third tonal unit (TU3); s8: TU1+TU2+40% of TU3; s9: TU1+TU2+60% of TU3; s10: TU1+TU2+80% of TU3; s11: TU1+TU2+TU3.

The mean confidence ratings increased progressively as a function of the amount of entering information. A pairwise comparison using the Neuman-Keuls test showed (at 0.01) the following set of results. The degree of confidence expressed for the answer to the last gate, which was very high (8.5/9), differed significantly from that given for all the other gates. In the same way, the degree of confidence expressed for gates s10 and s9 (respectively 6.1 and 5.6) differed significantly from that given for the gates s1 to s7. It should be noted that the degree of confidence for the answer to gate s7 was situated at the mid of the scale (4.4) and did not differ significantly from that given for the responses to gates s5 and s6.

To summarize, the analysis of the response patterns confirmed that all the participants gave systematic answers while hearing some series of stimuli. The analysis also showed that when a response changed during the series of stimuli, it changed in an irregular way mainly for A series and in a regular way for the Q series . The regularity testifies to the stabilization of the relevant answer while hearing a series. The listeners were sensitive to the lowered pitches, which precede the final rise, while being only fairly confident in the answers they gave.

DISCUSSION

The present study found its starting point in the linguistic analysis of intonational devices of total questions in French. It showed that, as soon as the composition of the intonation units exceeded two tonal units, a sequence of lowered pitches precedes the final contour, which according to question types can be rising or falling, (Di Cristo & Hirst, 1993). It was hypothesized that, in the formal marking of the interrogative mode by prosody, the lowered pitches may constitute a phonological property of yes-no question. The question was to verify the distinctive value of the declining pitch that precedes the final rise for ordinary listeners when they recognize yes-no questions. From a psychological point of view, to recognize an element of spoken language supposes the implementation of a process of comparison between two representations: a representation of the entry, which re-codes internally the information coming from the acoustic signal and a preliminary mental representation of the form to be recognized. So that to recognize the element concerned, the representation of the entry worked out by the listener must be discriminant with regard to the whole set of possible pairings. The gating technique that limits the amount of information input by controlling the temporal properties of the acoustic signal seemed a good way to determine when in the signal discriminant acoustic information becomes available to feed the process of comparison and thus lead to recognition.

The main result of this experiment is that recognizing the prosodic marking of sentence mode within the framework of the gating paradigm is difficult for listeners. Several reasons can be suggested. The recognition process is thought to require the following sequence of operations: (1) encoding the stimulus, (2) comparing this encoding with a preliminary internal representation and (3) formulating an appropriate response decision.

It should be noted first of all that the use of the gating paradigm could not contribute to the debate on the unsolved question of the existence of representations of pitch patterns in long-term memory. For the participants, performing the task was a question of comparing the representation of the fragments of utterances they heard (short and transitory) with internal representations strongly suggested by the response device that provided two possible written versions of a sentence. During silent reading, while sub-vocalizing, the participants could recreate the phonological and prosodic forms of the sentences (Bader, 1998; Fodor, 1998, 2002) to which the representation of the stimulus was then compared. The nature of the sentences (NP utterances) undoubtedly obstructed the convocation of a pitch pattern. The NP utterances selected as vectors of prosodic information in this experiment are easily interpreted as the statement of complete sentences (declrative or interrogative) if uttered in a precise communicative context. But within the experimental framework where they were presented, i.e. out of any communicative context, and in spite of the explicit indications given by the instruction, sometimes the sentence mode was not recognized exactly even though the prosodic information was completely available (9.4 % of the cases).

In addition, the operation of comparison concerned in the experiment requires engaging in a well-thought-out and conscious activity that, with regard to prosody, the ordinary circumstances of communication do not impose. During ordinary comprehension of spoken language, the amount of attentional resources allocated to the processes of intonation perception is generally considered to be less than the attentional resources allocated to the processes of lexical and syntactic encoding (Wales & Taylor, 1987). In this experiment the participants were confronted with a task that moved away from the ordinary mode of processing spoken language.

Moreover the recognition process not only requires more focal attention than usually devoted to prosody, but it requires making a subjective judgment on prosody. This involves decision-making processes which go beyond the decision-making process involved in ordinary spoken language processing. The gating paradigm is sometimes reproached for inducing extraneous mental operations such as guessing strategies intended to improve response accuracy. The participants are in particular constrained to guess when a stimulus does not yet contain discriminating information. In this experiment, in the absence of sufficient prosodic information, the listeners often considered a priori that the stimulus should come rather from an declarative statement (S) rather than from an interrogative one (Q). Questioned at the end of the experimental session, a number of them confirmed the strategic character of their responses at the beginning of a series of stimulus. The initial strategic answer supposed at the issue of the examination of the overall data, which showed more relevant answers for stimuli S that for the stimuli Q, was confirmed by the analysis of the individual response patterns and by the participants themselves.

Moreover the analysis of the individual patterns showed that another strategy coordinates itself with the initial strategic response. It consisted in awaiting the end of a series of stimuli to make a decision well informed by the properties of the last stimulus. Thus in 38.3 % of the cases, the listeners persevered in their initial answer ("S" or "Q") until the last stimulus was presented.

The analysis of the individual response patterns also showed that when the participants gave up their initial strategic answer in the course of hearing a series of stimuli, their answers were still often strategic. It was in particular the case (26.6 % of the patterns) when the responses changed from one stimulus to another until the end of a series (irregular pattern). But it was also the case when in a regular pattern a correct answer was given too early in the series to be allotted to discriminating characteristics of the stimulus. Many participants mentionned as a criterion for their response change the detection "of a rise". These accounts indicate that in addition to information derived from the sensory entry, the participants used information coming from their explicit knowledge (belief) that can be summarized as follows: a question is recognized by the fact that at one point "it flies away (ça s'envole)".

Thus, for certain series of stimuli, all listeners showed perseverations in their responses using the initial strategic response. But, also, all showed that they were, sometimes without success, attentive to information derived from the sensory input.

Considering now the assumption this study set out to check, the results showed two series of facts.

As with studies for languages other than French, our results confirm the decisive role of a final rise. The measurement of the time it takes to give a response of identification, which is usually not practiced with the gating technique, showed that listeners were slower to give their answer only for the first stimulus of each series. By the end of a series, the last but one and the last stimulus, the answers were given more quickly for stimuli S than for the stimuli Q. In S series the last two stimuli carried the falling pitch movement characteristic of the final contour of the declarative utterance. The decisive role of the final contour is also confirmed by a greater number of correct responses than for the other stimuli with regard to the last two stimuli of the S series on the one hand and the last stimulus of the Q series on the other hand. The analysis of the regular response patterns for the Q series makes it possible to specify these facts by showing the importance of the very end of the final contour only started in the last but one stimulus. At this point, only the correct answer was very close to 1 and was given with a very high confidence rating. In addition, about half of the regular patterns had a correct answer only for the last stimulus. This first series of facts was well summarized by a participant, which reported that s/he could answer towards the end, when on a certain word of the sentence "tout d'un coup ça s'envolait□" (suddenly it flew away).

The analysis of the regular patterns showed that part of the responses to the Q series (32.8 %) was organized according to a pattern compatible with the progressive discovery of the prosody of the input signal. The listeners who recognize a question from hearing the lowered pitches that precede the final rising movement were only fairly sure of the answer they gave. Thus, in spite of the fact largely asserted to have modified their initial strategic answer when they perceived a pitch rise, it appears clear that to identify the question certain listeners based their responses on the perception of the downstepping pitch that precede the final rise.

It remains however to consolidate this observation by other experiments using methodology which does not involve focal attention on the prosody or an explicit judgment with regard to it.

REFERENCES

Bader, J. D. (1998). Prosodic influences on reading syntactically ambiguous sentences. In F. F. J. D. Fodor (Ed.), Reanalysis in sentence processing. Dordrecht: The Netherlands Kluwer.

Cohen, J., Mac Whinney, B., Flatt, M., & Provost, J. (1993). An interactive graphical system for designing and controling experiments in the psychological laboratory using MacIntosh computers. Behavior Methods, Research, Instrumenys and Computer, 25, 257-271.

Cutler, A., Dahan, D., & Donselaar, W. (1997). Prosody in the comprehension of spoken language: A litterature review. Language and Speech, 40(2), 141,201.

Di Cristo, A. (1998). Intonation in French. In D. Hirst & A. D. Cristo (Eds.), Intonation systems : A survey of twenty languages (pp. 195-218). Cambridge: Cambridge University Press.

Di Cristo, A. (2000). Interpréter la prosodie. Paper presented at the XXIIIèmes Journées d'Etude sur la Parole, France, Aussois.

Di Cristo, A., & Hirst, D. J. (1993). Prosodic regularities in the surface structure of French questions. Paper presented at the European Speech Communication Association Workshop on Prosody, Lund.

Fodor, J. (1998). Learning to parse? Journal of Psycholinguistic Research, 27, 285-319.

Fodor, J. (2002). Psycholinguistics cannot escape prosody. Paper presented at the Speech Prosody Conference, Aix-en Provence.

Fon, Y.-J. J. (2002). A cross-linguistic study on syntactic and discourse boundary cues in spontaneous speech., The Ohio State University.

Grosjean, F. (1980a). How long is the sentence? Prediction and prosody in the on-line processing of language. Linguistics, 21, 501-529.

Grosjean, F. (1980b). Spoken word recognition processes and the gating paradigm. Perception and Psychophysics, 28, 267-283.

Grosjean, F. (1996). Gating. Language and Cognitive Processes, 11, 597-604.

Grosjean, F., & Hirt, C. (1996). Using prosody to predict the end of sentences in English and French: Normal and brain-damaged subjects. Language and Cognitive Process, 11(1/2), 107-134.

Hadding-Koch, K. (1961). Acoustico-phonetic studies in the intonation of southern Swedish. Lund: Gleerup.

Hadding-Koch, K., & Studdert-Kennedy, M. (1964). An experimental study of some intonation contours. Phonetica, 11, 175-185.

Hirst, D. J., & Di Cristo, A. (1984). French intonation: A parametric approach. Die Neueren Sprachen, 83, 554-569.

Hirst, D. J., & Di Cristo, A. (1998). A survey of intonation systems. In D. Hirst & A. DiCristo (Eds.), Intonation systems: A survey of twenty languages (pp. 1-44). Cambridge: Cambridge University Press.

Hirst, D. J., Di Cristo, A. & Espesser, R. (2000). Levels of representation and levels of analysis for the description of intonation systems. In M. Horne (Ed.) Prosody: Theory and Experiment (pp. 51-87). Dordrecht: Kluwer.

Le Goffic, P. (1993). Grammaire de la phrase française. Paris: Hachette.

Rossi, M. (1999). L'intonation, le système du Français: description et modélisation. Gap: OPHRYS.

Sˇaf árˇov á , M., & Swerts, M. (2004). On recognition of declarative questions in English. Proceedings of the Speech Prosody 2004 Conference, March 23-26, s.l., p. 313-316

Segui, J. F., L. (2000). Leçons de parole. Paris: O. Jacob.

Swerts, M., Bouwhuis, D., & Collier, R. (1994). Melodic cues to the percieved « finality » of utterances. Journal of Acoustical Society of America, 96(4), 2064-2075.

Vion, M., & Colas, A. (2002). La reconnaissance du pattern prosodique de la question : questions de méthode. Travaux Interdisciplinaires du Laboratoire Parole et Langage, 21, 83-106.

Wales, R., & Taylor, S. (1987). Intonation cues to questions and statements: how are they percieved? Language and Speech, 303, 199-211.

APPENDIX 1: Sentences

Sentences for the experimental trials:

les ragots du bureau *d'Amélie*

l'agonie du géant *d'Arabie*

le ronron du minet *du logis*

le ballon du gamin *d'Héléna*

le début du roman *de Zola*

le radeau du marin *d'Oléron*

les billets du magot *d'Odilon*

les manies du valet *du baron*

Sentences for the practice session:

les rabats du gilet *d'Ugolin*

l'enjeu des débats *des ruraux*

la magie du génie *d'Aladin*

APPENDIX 2: Instructions

You will hear truncated utterances (i.e. utterances which some words of the end are removed). The original utterances had varied intonations. They were STATEMENTS or QUESTIONS, as one meets some in the informal talk.

For example, you meet a former companion of college, which requires of you:

- What are you doing this year?
- I am studying at the Arts Faculty in Aix. (S)
- Studying at the Arts Faculty in Aix? And you like it (Q)?

Listen, you will hear it…

[Listening of the dialogue]

In the experiment, one took utterances like: "studying at the Arts Faculty in Aix", i.e. sentences without verb, and the last word group was removed. For this example here, “in Aix” removed. Which for this example gives the two following stimuli:

[Listening of the stimuli]

Your task will be to guess whether the whole utterance from which the stimulus results was an STATEMENT or a QUESTION. CAUTION! The stimuli which you will hear will be revealed gradually, i.e. you initially will hear: a first stimulus which comprises the words of the beginning of the utterance, for example: "studying at the Arts". Then a second stimulus that comprises the already heard words, plus a small bit of the word that follows. And so on until completely hearing the last word of the utterance.

With each time you hear a stimulus, you will have to say from which utterance, A or Q, this stimulus comes from. For that, you should press the keys labeled "A" and "Q" on the keyboard.

As you cannot always be certain of your response, you should also express the degree of confidence, which you accord to your answer. You should express your confidence by means of the keys "1" to "9" on the number pad where: "1" will mean "not certain at all ", "9" will mean "very certain ".

Also note: at the end of the experiment you will be asked what were the CUES which enabled you to decide that it was a stimulus coming from a STATEMENT or from a QUESTION.

First we will carry out some practice trials so that you will be familiarized with the organization of the experiment, then one will move on to the experiment itself.

You have all your time to answer.

Press the space bar to begin